Long-term effects of two universal interventions on adolescent delinquency, substance use, and externalising behavior

Averdijk, M ; Zirk-Sadowski, J ; Ribeaud, Denis ; Eisner, Manuel

DOI: https://doi.org/10.1007/s11292-015-9249-4

Running head: LONG-TERM EFFECTS

Long-term effects of two childhood psychosocial interventions on adolescent delinquency, substance use, and anti-social behavior: A cluster randomized controlled trial

Margit Averdijk[1], Jan Zirk-Sadowski[23], Denis Ribeaud[1], & Manuel Eisner[2]

[1] Criminological Research Unit, Department of Sociology, Swiss Federal Institute of Technology Zurich (ETH)

[2] Violence Research Centre, Institute of Criminology, University of Cambridge

[3] Eurotitis-2 Study Group, Department of Psychology, University of Cambridge

Author Note

Margit Averdijk, Department of Sociology, Swiss Federal Institute of Technology Zurich (ETH), Switzerland; Jan Zirk-Sadowski, Institute of Criminology and Department of Psychology, University of Cambridge, UK; Denis Ribeaud, Department of Sociology, Swiss Federal Institute of Technology Zurich (ETH), Switzerland; Manuel Eisner, Institute of Criminology, University of Cambridge, UK.

The research reported in this manuscript was financially supported by the Swiss National Science Foundation, the Jacobs Foundation, the Swiss Federal Office of Public Health, the Canton of Zurich Ministry of Education, and the Julius Baer Foundation. The authors would like to express their sincere thanks to the youths, parents, and teachers for participating in the study. Moreover, the authors are grateful to all the interviewers and undergraduate students for their help in data collection and coding.

Correspondence concerning this article should be addressed to Margit Averdijk, ETH Zurich, Department of Sociology/Criminological Research Unit, Weinbergstrasse 109, 8092 Zurich, Switzerland. E-mail: margit.averdijk@soz.gess.ethz.ch. Tel: +41 44 632 97 31 Fax: +41 44 632 13 41

Abstract

Objectives: Examine the long-term effects of two childhood universal prevention programs on adolescent delinquency, substance use, and antisocial behavior.

Methods: The cluster randomized controlled trial involved 56 schools and 1,675 children in Zurich, Switzerland. Two evidence-based interventions, namely the social-emotional skills program PATHS and the cognitive-behavioral parenting program Triple P were implemented during the first two years of primary school, at ages 7 and 8. Outcomes were measured at ages 13 and 15 and included youth self-reports and teacher assessments. Multilevel models were used to account for the clustered nature of the interventions. Effects were estimated with inclusion of baseline covariates.

Results: Across 13 outcomes related to delinquency, substance use, and antisocial behavior at ages 13 and 15, only two non-negligible effects were found. The first was a reduced prevalence of police contacts in the PATHS condition (ES = -.225). The second was a difference in competent conflict resolution skills in the combined PATHS + Triple P condition compared to the context (ES = .259), but in the unexpected direction: Participants in the combined treatment appeared to be less competent than their control group peers. All other effects were either statistically non-significant or negligible in size (i.e., ES < |.200|).

Conclusions: Even 'evidence-based' interventions may have few long-term effects on delinquency, substance use, and antisocial behavior. Our findings add to the small literature on the long-term effectiveness of early universal prevention in field settings.

Keywords: Developmental prevention, juvenile delinquency, cluster-randomized controlled trial, social skills program, parent training

Long-term effects of two childhood psychosocial interventions on adolescent offending, substance use, and antisocial behavior: A cluster randomized controlled trial

There is increasing evidence that preventive action aimed at reducing risk factors and promoting protective factors during infancy and childhood can reduce the likelihood of later crime (Farrington and Welsh 2003; Kaminski et al. 2008; Piquero et al. 2009). Among the most promising approaches are programs that target parenting skills and/or promote social and emotional skills during childhood. However, two pressing concerns relate to the lack of consolidated knowledge on the long-term effectiveness of early prevention strategies and the difficulty of replicating positive effects in field trials outside the context of the original research. The present analysis contributes to this literature by examining the results of a randomized controlled trial (RCT) that, independently of program developers, investigated the effectiveness of two universal interventions in a field setting in Zurich, Switzerland. The programs were implemented during the first two years of primary school when the children were aged 7 and 8. We examined intervention effects on delinquency, substance use, and antisocial behavior when the participants were aged 13 and 15, using a multi-informant approach.

The past decade has seen a significant increase in the number of experimental and quasi-experimental studies examining the effects of parent training and social skills programs on conduct problems during childhood. Several meta-analyses and systematic reviews have summarized the current state of the evidence (e.g., Beelmann and Lösel 2006; Durlak et al. 2011; Furlong et al. 2012; Piquero et al. 2009, Farrington and Welsh 2003; Wilson et al., 2012). Overall, the findings have consistently shown that, on average, both parent training and social skills programs reduce child behavior problems, although effects sizes vary significantly between studies, target groups, and geographic locations. Programs are especially effective if they have been implemented with fidelity and a dosage that is commensurate to the severity of

the problem in the target population. Meta-analyses have also revealed larger effect sizes in studies with smaller sample sizes (Farrington and Welsh 2003; Koehler et al. 2012), shorter follow-up periods (e.g., Dekovic et al. 2011; Lösel and Beelman 2003), and high-risk samples (e.g., Koehler et al. 2012), and in studies carried out by program developers (Beelmann and Lösel 2006; Wilson et al. 2012). Thus, a recent meta-analysis concluded that there is "an obvious need for more effectiveness studies, carried out by independent researchers" (Smedler et al. 2014).

In addition, very few methodologically rigorous studies have examined the long-term effectiveness of parent training and social skills programs on delinquency and externalizing behavior (Piquero et al. 2009; Smedler et al. 2014). In a recent review, Farrington and Welsh (2013) found only 12 studies that had followed participants in the long-term (at least 10 years), had sufficient sample size (at least 50 persons per condition), and included offending as an outcome. Most of these programs targeted high-risk populations and the majority were demonstration projects in the sense that program implementation and study design were coordinated by the program developers. In contrast, none of the included studies examined the long-term effects of universal preventive measures implemented during childhood on delinquency during adolescence.

The interventions

Two internationally known and evidence-based universal interventions were implemented, namely the social and emotional skills program Promoting Alternative Thinking Strategies (PATHS) and the parenting program Triple P. They were selected on the basis of an extensive city-wide needs and resources assessment conducted in 2001/2 and the strong international empirical evidence suggesting reductions in child problem behavior following the interventions (Eisner et al. 2011). As parenting, social-emotional skills, and early problem behavior have repeatedly been shown to be among the best developmental predictors for later delinquency

and problem behavior (Eisner and Malti 2015; Farrington 2012), programs that target these risk factors have considerable potential to reduce antisocial behavior beyond childhood into adolescence. Thus, the idea of a risk-focused prevention strategy as well as a developmental perspective on juvenile crime guided the selection of the two interventions (Eisner and Ribeaud 1995). For Triple P, ours is the first study worldwide that examines long-term effects on offending. For PATHS, the present study is the only study other than the Fast Track Project (Conduct Problems Prevention Research Group 2010) that reports long-term effects on offending.

Both interventions were implemented as universal, population-level programs, i.e., as interventions targeted at all children rather than a particular sub-group of children at risk. The rationale for this was threefold. First, a considerable percentage of children and adolescents in Switzerland is exposed to the risk of displaying problem behaviors. Self-reported rates of anti-social and risk behaviors in Switzerland are average or above compared to other European countries and the U.S. (Craig et al., 2009; Currie et al., 2012; Simons-Morton et al., 2009). For example, among 15-year-olds, 13% of girls and 26% of boys reported weekly alcohol use, 11% of girls and 17% of boys reported cannabis use in the last 30 days, and 4% of girls and 16% of boys reported physical fighting. Bullying perpetration was reported by 11% of girls and 25% of boys (Currie et al., 2012). According to the International Self-Report Delinquency (ISRD) study, the past-year prevalence of offending among 13- to 16-year-olds in Switzerland was 13% for frequent violent offenses, 3% for rare violent offenses, 8% for vandalism, 9% for shoplifting, 5% for rare property offenses, 5% for computer hacking, and 3% for drug dealing (Killias et al., 2010). These rates are average or above-average compared to other Western countries participating in the ISRD (Enzmann et al, 2010). Thus, our study aimed to address these risk behaviors in the normative population. Second, universal, school-based prevention and parenting programs have been found to be effective among normative populations as RCTs

have shown reductions in problem behavior, substance use, and crime in the intervention participants relative to the control group in childhood and adolescence (e.g., Bradley et al., 2003; Dishion et al., 2002; Lösel et al., 2013; Spoth et al., 2002). There is evidence that these effects may be durable and meaningful. For example, in one trial evaluating a program teaching general life skills and skills for resisting social influences to use drugs, there were 44% fewer drugs users and 66% fewer polydrug users among participants relative to controls 6 years after baseline (Botvin et al., 1995). Third, although the city of Zurich was found to have a good and dense system of selective interventions, there was a lack of universal provision that would serve all children in a cohort (Eisner et al., 2003). Thus, there was an unmet demand for universal programs, which resulted in the implementation of the two said programs as well as a test of their effectiveness.

PATHS is a research-based program to enhance affective, cognitive, and social competencies and reduce aggression and behavior problems in primary school-aged children (Kusché and Greenberg 1994; Domitrovich et al. 2007; Greenberg and Kusché 1998; Riggs et al. 2006). It places particular importance on emotional learning processes, which are assumed to precede cognitive learning. The program uses a variety of strategies to increase emotional skills, self-regulation, and the integration of various developmentally relevant competencies. It has been thoroughly evaluated (Crean and Johnson 2013; Kam et al. 2003; Kam et al. 2004) and is one of the programs listed as effective by the Blueprints for Healthy Youth Development at the University of Colorado Boulder (Greenberg and Kusché 2002). However, most previous evaluations have been conducted by the program developers.

Triple P is a multi-level parent training program based on cognitive-behavioral therapy (Sanders, 1999, 2012). It comprises five levels of intervention, ranging from universal parent information to a specialized intensive intervention for families with different sources of distress (see Nowak and Heinrichs 2008). Triple P is among the most thoroughly evaluated parent

training programs worldwide. A recent meta-analysis conducted by the program developers examined findings from 101 Triple P studies (Sanders et al. 2014) and reported positive effects, mainly based on parent assessments, over a broad range of outcomes, including parenting practices and children's social, emotional, and behavioral outcomes. However, another meta-analysis highlighted methodological problems in many relevant studies as well as a lack of transparency regarding the disclosure of financial conflicts of interests (Wilson et al. 2012). Also, no intervention effects were found according to teacher assessments.

Previously published results of the Zurich trial

While no prior study has investigated long-term treatment effects into adolescence in the Zurich trial, four studies have reported on the effects of the interventions up to age 11. Eisner et al. (2007) reported effects up to 6 months (for PATHS) and one and a half years (for Triple P) after the intervention. Analyses suggested some small effects for Triple P, indicating greater reductions in parent-reported corporal punishment, impulsive parenting techniques, and improved family climate, but also increased teacher-reported non-aggressive conduct problems in the intervention group compared to the control group. For PATHS, two small effects were found, namely greater reductions in self-reported aggressive problem solving and greater teacher-assessed emotion recognition in the intervention group compared to the control condition.

Malti et al. (2011) examined the effects of both interventions on children's aggressive behavior, Attention Deficit Hyperactivity Disorder (ADHD), non-aggressive conduct disorder, prosocial behavior, and social-cognitive skills up to age 11. A multi-informant design was used with the children, parents, and teachers as the raters. The study used an intention-to-treat design and growth-curve models to assess the effects of the programs on children's externalizing behavior and social competence over time. Its findings showed small effects in the PATHS-only condition on reduced teacher-rated externalizing behavior (aggression and ADHD) and

parent-rated aggression, but the majority of outcomes was unaffected and no positive effects could be found in the combined Triple P and PATHS condition. In addition, there was no evidence that Triple P by itself decreased externalizing behavior or promoted social competence.

In a subsequent study, Eisner et al. (2012) examined the effects of Triple P among highly adherent parents who had attended all four Triple P training sessions. Using a propensity score matching approach for modeling self-selection into treatment, the parents who received all training sessions were compared to the untreated parents in the control condition. The groups were matched on 54 baseline characteristics. The results revealed that there were no effects on five dimensions of parenting practices (parental involvement, positive parenting, poor monitoring, erratic discipline, and corporal punishment) or five dimensions of child problem behavior, assessed by the child, the parent, and the teacher (prosocial behavior, internalizing problems, ADHD, non-aggressive conduct problems, and aggressive behavior) up to age 11.

In a fourth publication, Malti et al. (2012) examined whether the effects of PATHS on teacher-rated aggressive behavior, ADHD, and prosocial behavior at follow-up were moderated by child baseline characteristics or the economic situation of the family. Developmental variables moderated the effect of PATHS on ADHD and aggression at follow-up: Children who received the intervention and displayed higher levels of moral emotions and/or competent problem-solving showed greater decreases in aggression and ADHD than children with low moral emotions and/or less competent problem-solving techniques. However, most outcomes were unaffected by the main effects or moderation of treatment effects.

In sum, the extant analyses have revealed limited effects of the interventions on externalizing behavior problems until age 11. The present paper extends these analyses by examining not only long-term effects, but by specifically focusing on treatment effects on

delinquency and substance use, i.e., two behavior dimensions that typically show increases during early adolescence (e.g., Mason et al. 2003; Wills et al. 2001; Windle and Wiesner 2004). It also contributes to the limited literature on possible sleeper effects, where treatment effects only emerge with some delay after the intervention (e.g., Vitaro et al. 2001).

Method

The data were derived from the longitudinal Zurich Project on the Social Development of Children and Youths (z-proso; Eisner et al. 2011). Its target population consisted of all 2,520 children who entered year 1 of public primary school in 2004 in the city of Zurich, Switzerland. Zurich is the largest city of Switzerland with a population of approx. 420,000. It is highly affluent, has a low unemployment rate, and a high proportion of socially disadvantaged immigrant minorities. The vast majority of children attend state schools, which have a very good reputation.

Sampling was based on a cluster randomized approach with schools, rather than individuals or classes, as the randomization units in order to minimize possible contamination or spillover effects between treatment conditions. The sampling frame was all 90 public primary schools in Zurich. The schools were first blocked by school size and socio-economic background of the catchment area. Then, a stratified sample of 56 schools was drawn, comprising 1,675 children. All selected schools participated in the study. Due to the stratified sampling procedure, the 56 schools formed 14 “quadruplets” of similar size and socio-economic background. Schools in disadvantaged school districts were slightly overrepresented in the sample.

The present analyses used data from the baseline assessment in wave 1 (collected in 2004/2005) and follow-up data from waves 5 (2011) and 6 (2013). Wave 1 included data from the primary caregiver, the child, and the teacher. In waves 5 and 6, the parent assessments were discontinued and only child and teacher assessments were collected.

In wave 1, when the mean age at the time of the child interview was 7.45 years, the response rate was 81% for the children (N = 1,361), 74% for the parents (N = 1,240), and 81% for teacher assessments (N = 1,350). Among those children who participated in the child interview at wave 1, 1,158 (85%) and 1,248 (86%) took part in waves 5 (mean age 13.7 years) and 6 (mean age 15.4 years). Among the teachers who participated at wave 1, 1,066 (84.0%) and 1,092 (84.5%) participated in waves 5 and 6. Details are presented in Figure 1.

The sample was 52% male. In 46% of all cases both parents were born outside of Switzerland. Compared to Swiss parents, non-Swiss parents had a lower level of education, higher rates of unemployment, were less integrated into their neighborhood, and participated less in their children's school (Eisner et al. 2008).

Interventions

A 2x2 factorial design was used to examine whether the isolated interventions as well as the combination of the parenting program and the child skills training yielded effects in preventing problem behavior (Beelmann and Lösel 2006; Farrington and Welsh 2003). The schools in each quadruplet were randomly allocated to four treatment conditions: PATHS only, Triple P only, PATHS + Triple P combined, and control group. In order to achieve optimal implementation quality, the programs were delivered sequentially: Triple P was implemented between waves 1 and 2 (i.e., year 1 of primary school), whereas PATHS was implemented between waves 2 and 3 (i.e., year 2). The sample sizes in the groups were n = 442 for PATHS only, n = 422 for Triple P only, n = 397 for PATHS + Triple P, and n = 414 for the control group.

Substantial efforts were made to achieve high implementation quality (Eisner et al. 2009; Malti et al. 2011). The developers of both programs were consulted about criteria for high quality implementation, but were neither involved in the intervention delivery nor in the data

analysis. Thus, the delivery of the programs was separated from the data collection and data analysis.

PATHS

The present study used the PATHS version that was also used in the second school year of the Fast Track Project (Bierman et al. 2010). It is a 1-year program that includes 46 lessons. The program was adapted to the Swiss school system and the materials were tested in a pilot study (see Eisner et al. 2006). The lessons addressed problem-solving skills, social relationships, self-regulation, rule understanding, emotion understanding, and positive self-esteem. Teachers received two days of training. Five coaches were trained and supervised by an experienced Dutch expert, who has led the PATHS training for teachers in the Netherlands since the early 1990s. The coaches visited each class four to six times throughout the implementation period. Additional measures to support implementation included a refresher seminar held at midterm, a PATHS newsletter, and coach-led group meetings where issues related to the program could be discussed. For the purpose of the evaluation study, the city of Zurich had made the PATHS curriculum compulsory for teachers in the intervention group, meaning that there was no self-selection into the treatment condition.

Data collected to monitor implementation included teacher and child questionnaires as well as classroom observations (Greenberg and Kusché 2002). Indicators suggest a high overall implementation quality. Teachers taught PATHS an average of 67 minutes per week, with an average of 2.4 sessions per week. The coaches assessed a total of 308 PATHS lessons. Depending on the assessed aspect (e.g., child motivation, quality of classroom management, cooperation with coaches), between 74 and 81% of the lessons were rated as high quality.

Triple P

Level 4 Triple P was implemented. Its core element was a group-based course with four units of 2 to 2.5 hours. The units addressed themes such as positive parenting, techniques to

support desired behaviors, and routines to help avoid the escalation of conflicts. Additionally, the program included up to four follow-up telephone contacts with each participant. Participation was free of costs and travel distances were generally below one mile. The program was offered on different weekdays and times of the day, and a free child-care service was available. Additional efforts were made to recruit families with an immigrant background; besides German, the program was offered in Albanian, Portuguese, and Turkish, which are spoken by sizeable communities of immigrants in Zurich (Eisner and Ribeaud 2007). Courses were delivered by licensed facilitators selected in collaboration with Triple P Switzerland. Program delivery was coordinated by the Municipality of Zurich.

The implementation team organized 41 Triple P courses, of which 33 were held in German, three in Turkish, two each in Portuguese and Albanian, and one in English. In August 2004, when the children entered primary school, all parents were informed about the study and invited to participate in the study and the planned prevention programs. On the first parents' evening, Triple P trainers gave a 10- to 15-minute presentation about the training in the target classes. In classes with many immigrant parents, bilingual trainers provided translations. In October 2004, the parents received written information packages and registration talons, both of which were translated into all course languages. Immigrant parents received the materials in two languages. Courses were delivered between May and July 2005. The parents of 27.0% of the children attended at least one session; the parents of 2.2% of the children attended only one session, the parents of 1.7% of the children attended only two, the parents of 4.3% of the children attended three, and the parents of 18.8% of the children completed all four. This is comparable to or better than other studies, which have shown initial participation rates of 20 to 30% of the target population and considerable attrition as the program progresses (e.g., Heinrichs et al., 2005; McTaggart and Sanders, 2003; Webster-Stratton et al., 2001). About 60% of those who had attended the training reported using a significant part of the learned

techniques 3 months after the intervention. Parents who completed the program were more likely to come from breadwinner families, to be Swiss, to have high socio-economic status, and to be highly integrated in neighborhood social networks. However, compliers did not differ from non-compliers with regards to parenting problems or child problem behavior (Eisner and Meidert 2011).

The program was delivered to high standards. Participant overall satisfaction with the program was 4.33 (*SD* = 0.89) and provider competency was rated at 4.65 (*SD* = 0.73) on a five-point scale. Course providers estimated that 93% of the full course material was delivered during the sessions.

Outcome measures

Self-reported delinquency. The youths were administered 15 delinquency items. Items included the past-year prevalence of running away from home, stealing at home, stealing at school, shoplifting something worth more than 50 dollars, shoplifting something worth less than 50 dollars, vehicle theft, driving without a license, burglary and stealing from a car, drug dealing, graffiting, vandalism, carrying a weapon, threatening and extortion, robbery, and assault. In total, 42.3% of youths reported to have committed at least one type of delinquency at age 13 and 49.8% at age 15. A total variety scale that represented the number of delinquency types was computed (Bendixen et al. 2003). Variety scales have been termed 'the preferred criminal offending scale' as they display high reliability and validity, are less skewed than frequency measures, and are not compromised by high-frequency crime-types of low seriousness (Sweeten, 2012).

Teacher-reported deviance. For the teachers, 7 items measured deviance, including truancy, assault with injury, carrying a knife or weapon, using threats to get something, smoking cigarettes, drinking alcohol, and taking illegal drugs in the past 6 months. According

to the teachers, 14.1% of the youths committed at least one type of deviance at age 13, while 32.1% did so at age 15. We constructed a variety score.

Self-reported police contacts related to an offense. For youth-reported delinquency, follow-up items assessed the self-reported prevalence of a police contact due to each type of delinquency. In total, 7.4% of the youths reported a police contact at age 13 and 10.2% did so at age 15. We constructed an overall prevalence score across all items. Unfortunately, we did not have official data to offer an alternative measure of police contacts or validate the self-report measure, which is a limitation of our study. Furthermore, given that the police contact measure was assessed as a follow-up question to the delinquency items, it is possible that youths who under-reported delinquent behavior were missed with respect to police contact.

Substance use. For the youths, 4 items measured the past-year consumption of tobacco, alcohol, strong liquor, and marijuana. Answers were given on a 5-point scale from 1 (‘never’) to 5 (‘daily’). Of the youths, 39.6% reported that they had consumed at least one of the mentioned substances at age 13, while 75.7% had done so at age 15. After recoding the category 1 to ‘0’ and the categories 2 through 5 to ‘1’, we computed the final score as a variety scale.

For the teachers, we computed a variety scale of 3 items, which were a sub-sample of the teacher-reported deviance measure and assessed the prevalence (0 ‘no’ and 1 ‘yes’) of smoking cigarettes, drinking alcohol, and taking illegal drugs in the past 6 months. According to the teachers, 3.8% of the youths had consumed at least one of the mentioned substances at age 13, while 17.3% had done so at age 15. Thus, the final score computed the sum of these items.

Aggressive behavior. Aggressive behavior was measured on the Social Behavior Questionnaire (SBQ). Tremblay et al. (1991) reported internal consistency, test-retest reliability, stability over time, and concurrent and predictive validity of the SBQ. We used both youth- and teacher-reported aggressive behavior. Slightly different versions of the instrument were administered to the teachers and the youths (e.g., “threatens others”). The youth measure

was constructed as a 12-month mean score across 9 items (3 proactive, 3 reactive, and 3 physical aggression items). The teacher measure was constructed as a 6-month mean score across 11 items (4 proactive, 3 reactive, and 4 physical aggression items). Responses were recorded on a 5-point Likert scale. The reliabilities were α = .84 (age 13) and α = .83 for the youths and α = .93 (age 13) and α = .92 (age 15) for the teachers.

Self-reported peer aggression. Peer aggression is conceptually related to, but not the same as bullying perpetration; while bullying is aggressive behavior, repeated, and characterized by an imbalance of power (Olweus 1999), peer aggression focuses more generally on aggressive acts between children and youths. The self-report peer aggression scale was derived from Olweus (1993) and included four types: Teasing in a mean and hurtful way, stealing and destroying possessions, physical violence, and rejection/exclusion in the past 12 months (α = .78 at age 13, and α = .75 at age 15). We computed the mean.

Prosocial behavior. The youths and the teachers assessed prosocial behavior using the SBQ (e.g., "shares things with peers"). The youth measure was constructed as a 12-month mean score across 8 items and the teacher measure as a 6-month mean score across 7 items. The reliabilities were α = .82 (age 13) and α = .80 (age 15) for the youths and α = .93 (age 13) and α = .90 (age 15) for the teachers.

Conflict resolution. An 8-item conflict resolution scale was adapted from Wetzels et al. (2001) and administered to the youths. Four of the items were used to create a measure for aggressive conflict resolution strategies and four for competent conflict resolution strategies. The reliabilities were α = .72 (age 13) and α = .67 (age 15) for the aggressive strategies and α = .71 (ages 13 and 15) for the competent strategies.

Non-aggressive conduct disorder (NACD). NACD was assessed by the teachers using the SBQ (e.g., "tells lies and cheats"). It included 6 items (2 for oppositional defiant disorder and

4 for non-aggressive conduct disorder). The reliabilities were $\alpha = .83$ (age 13) and $\alpha = .85$ (age 15). We computed the mean.

Baseline control variables. In RCTs, covariates are included for two reasons. The first is that baseline measures that reflect stable between-individual differences in the relevant outcome dimension increase the statistical power of the analyses because they model some proportion of the variation of the outcome measure. Thus, including a baseline measure that is expected to co-vary with the outcome results in a greater chance of detecting effects (e.g., Cohen et al. 2013). The second occurs in situations where baseline balance on some relevant variables could not be achieved (Roberts and Torgerson 1999).

To account for between-individual variation in outcome measures at baseline, we included wave 1 parent, teacher, and child measures of externalizing behavior. Furthermore, we included gender (52% were males), attendance of a special needs class (91% were in a regular class), non-Swiss nationality (55% of participants had at least one Swiss parent), and socioeconomic status (SES) to control for basic demographic patterns in the data. SES was based on coding the caregiver's current profession. The codes were then transformed into an International Socio-Economic Index of occupational status (ISEI) score (Ganzeboom et al., 1992). We took the highest ISEI score of the two caregivers across waves 1 and 5/6 ($M = 45.83$, $SD = 18.72$).[1]

Statistical Analysis

We used multilevel modeling to account for similarities between children in the same schools (Muthén and Muthén 1998-2012). Random intercept multilevel models were used to assess the effects of PATHS and Triple P. The treatment condition coding was consistent with a factorial approach (see Jaccard and Turrisi 2003; Montgomery et al. 2003). The variable indicating membership of the PATHS condition was coded "1" for PATHS, "1" for PATHS + Triple P, "0" for the control condition, and "0" for Triple P. The variable indicating

[1] At wave 6, the ISEI score was collected only among those youths who did not participate at wave 5.

membership of the Triple P condition was coded "1" for Triple P, "1" for PATHS + Triple P, "0" for the control condition, and "0" for PATHS. In a second analytic step, these two binary variables were multiplied into an interaction term (PATHS x Triple P) in order to examine whether the combined interventions yielded effects over and above those of the isolated interventions. This was done because the study was designed as a 2x2 factorial design. The cross-product of the PATHS + Triple P intervention examined whether adding PATHS improved the effect of Triple P and vice versa (Malti et al, 2011). The models incorporated two levels: Youths (Level 1) and school at treatment (Level 2).

Linear models estimated fixed effects with the robust maximum-likelihood method that handles skewness of the data (Muthén & Muthén 1998-2012). For youth-reported substance use and police contacts, where the distributions included only integer values, ordered-logistic models were used. Significant treatment effects were considered to be non-negligible if the Cohen's *d* effect size (ES) was by convention at least small (i.e., ES > |.200|) (Rice and Harris 2005).

For the teachers, 26% of the data points in the outcome measures were missing and for the youth reports 17%. Missing data were handled with robust full-information maximum-likelihood estimation (FIML) including means and variances of continuous predictors in the model, assuming existence of their latent factors to avoid loss of information (Muthén and Muthén 1998-2012). FIML (e.g., Enders and Bandalos 2001; Enders 2001) is the preferred method for handling missing data patterns, providing unbiased and more efficient estimates than alternative methods including listwise deletion and regression or mean imputation. It is superior to multiple imputation in multilevel models (Larsen 2011). FIML resembles obtaining predicted scores for a missing data pattern by regressing one measure on the remaining measures used in the analysis. Instead of imputing missing data, FIML estimates the model parameters and their standard errors by using all available information on observations for a

given parameter. This means that for one beta coefficient, the maximized probability of its value can be generated based on a different sample size than for another beta coefficient included in the same model.

We ran two series of analyses. The first used the dataset with all available data points of the target sample (n = 1,675). Due to missing data on both the outcome and predictor variables, these analyses effectively used data from n = 1,580 participants. The second used stricter inclusion criteria, including only participants who participated both at age 7 and at either age 13 or age 15 (n = 1,275). This was done for the purpose of sensitivity analysis in order to examine whether non-participation and attrition affected the results. To examine the possible robustness of the results, we report on both types of analyses below, i.e., first for the 1,580 participants and next for the restricted dataset with 1,275 participants.

Results

No statistically significant baseline differences on any of the teacher, parent, or child measures of externalizing behavior were found (see Malti et al. 2011). Initial descriptive analyses were conducted to examine levels of delinquency, substance use[2], and antisocial behavior at ages 13 and 15. Table 1 reports the means and standard errors by time (left panel), and treatment condition by time (right panel). Results showed that significant proportions of adolescents reported delinquent activities (see also the descriptive statistics mentioned in the Method section).

Inspection of intra-class correlations

We estimated preliminary unconditional two-level models for ages 13 and 15 separately to inspect intra-class correlations (ICCs). ICCs indicate the proportion of variance in the outcome variables that can be attributed to the school level. At age 13, the median ICC at the school

[2] We conducted additional analyses for alcohol use separately and found no significant effects.

level was low at .02 (range .01 to .03 across outcomes). Across youth-reported measures at this age the median ICC was .01. Across teacher-reported measures it was .03.

At age 15, the median ICC at the school level was also low at .01 (range .01 to .05 across outcomes). Across youth-reported measures at this age the median ICC was .01. Across teacher-reported measures it was .03. These findings suggest that only a marginal proportion of the total variation in the outcome variables was due to school characteristics.

Intervention effects

Next, multilevel analyses were estimated to assess treatment effects on the outcome variables. The results for the full sample are reported in Table 2. There was no evidence for non-negligible treatment effects at age 13 or 15. Although there were several statistically significant effects for PATHS and Triple P at both ages, none exceeded ES < |.200|. At age 13, the statistically significant effects referred to a lower prevalence of police contacts in both the PATHS (ES = -.157) and the Triple P (ES = -.137) conditions compared to the youths in the control group. At age 15, no statistically significant effect emerged for PATHS, whereas youths in the Triple P condition showed a greater reduction in teacher-reported deviance (ES = -.109) and aggressive behavior (ES = -.149) than their control group counterparts. However, youths in the Triple P condition also showed *lower* competent conflict resolution than the control group (ES = -.100), although this effect, too, remained negligible.

Next, the analyses were estimated on the restricted dataset in which only respondents who participated at both age 7 and at either age 13 or age 15 were included (n = 1,275). The results are displayed in Table 3. These analyses revealed one non-negligible effect for PATHS, indicating that youths in the PATHS condition were less likely to report a police contact at age 13 than youths in the control condition (ES = -.225). The other statistically significant but negligible effect for PATHS was opposite to hypothesized, as it indicated a greater reduction in competent conflict resolution skills in the treated youths compared to the control group (ES

= -.121). For Triple P, there were no non-negligible effects. The largest effect sizes indicated that the treated youths showed lower teacher-reported deviance (ES = -.138) and aggressive behavior (ES = -.135) at age 15 and less youth-reported police contacts (ES = -.123) at age 13 than the others. Contrary to expected, the treated youths in the Triple P condition also showed lower prosocial behavior (ES = -.104) and less competent problem solving (ES = -.117) compared to their control group counterparts. However, these effects remained negligible.

Combined intervention effects

Our final set of analyses assessed whether the combined treatment of PATHS and Triple P yielded effects. To this end, we included an interaction term (PATHS x Triple P) into the models. The first series of analyses was run on the full dataset (n = 1,580). No statistically significant effects were found. The second series of analyses was estimated on the restricted dataset (n = 1,275). In this case, a statistically significant and non-negligible effect was found for competent conflict resolution skills at age 13 (ES = .259; Table 4). Paradoxically, the combined intervention in comparison to the control group led to markedly worsened skills, and did not differ in this respect from the isolated PATHS and Triple P interventions (see Figure 2).

Discussion

The purpose of this study was to test the long-term effectiveness of the PATHS and Triple P interventions implemented at ages 7 and 8 on delinquency, substance use, and antisocial behavior at ages 13 and 15. The study was based on a large scale clustered randomized controlled trial carried out in Zurich, Switzerland. Although the two tested interventions have been considered by their originators to be reliable means for reducing later antisocial behavior (Greenberg et al. 1998; Sanders 1999, 2012), our work was the first to assess such long-term effects independent of program developers. Thus, one strength of this trial was that the interventions were delivered in a large field trial to high standards by organizations that were

independent of the research team and the program developers. Also, no actors with a financial interest in positive outcomes were involved in either study design or data analysis. This is important because treatment effects of psycho-social interventions found in studies with a conflict of interests often cannot be replicated in independent trials (Eisner 2009).

The present study found practically no evidence for long-term intervention effects. Although some of the found effects reached statistical significance, most of these remained marginal in terms of their effect size. In total, 52 effects were reported, of which only two were found to be non-negligible. The first was a lower prevalence of police contacts reported by the youths in the PATHS condition compared to the other youths at age 13. Although this effect indicated a desirable intervention effect, it was isolated in the sense that it did not appear at age 15 and was found only in the analyses on the restricted dataset and not on the full data set, meaning that the effect may not be generalizable. Furthermore, the interpretation of this finding is complicated by the absence of intervention effects on self-reported delinquency. The second was *less* competent conflict resolution in the youths in the combined PATHS + Triple P condition, which was contrary to expected. At the same time, we found no non-negligible effects of Triple P alone. This is partly in line with the only other study on long-term effects of Triple P, namely the trial conducted by Heinrichs et al. (2014) in Brunswick, Germany. That study found no long-term effects of Triple P on child behavior problems across four years. Thus, it appears that the PATHS social and emotional skills program emerges from our study as a slightly more promising intervention than the cognitive-behavioral Triple P program.

Besides these results, we found some statistically significant, but negligible effects, some of which were in the opposite from expected direction, namely lower competent problem solving and less prosocial behavior in the youths who received the (isolated) interventions compared to the control group. Although these effects remained negligible, future studies that include mediator analyses to examine potential explanations would be helpful. In addition, future

analyses on the role of family environments, deviant peer involvement, and cascading effects of early problem behavior on later delinquency and substance use may provide insights into the question of why only few intervention effects were found. Finally, future analyses may also examine whether any of the effects were associated with effective participation in the treatment (in the case of Triple P) or with higher quality delivery (in the case of PATHS). Although our previous analyses on highly adherent parents were not favorable (Eisner et al., 2012), further analyses may use alternative methods to analyze compliance, such as the complier average causal effect estimation approach (Jo, 2002).

There may be several explanations for the overall lack of long-term effects in this trial. First, most interventions with demonstrated long-term positive effects on delinquency to date are relatively intensive programs that target children with a combination of risk factors, whereas ours consisted of two universal programs. It may be that universal early interventions are not powerful enough to obtain long-term effects, and that they would need to be maintained over a longer period of time and/or be combined with more intensive interventions for high-risk groups. One exception to this literature, however, is the study by Lösel et al. (2013), which examined the long-term effects of a social skills program and a parent training program in Germany. After 5 and 10 years, several desirable program effects were found, although these were mainly small with statistical significance becoming rare over time. There were several ways in which this study differed from ours, which may explain why the study findings were different. The interventions in the study by Lösel et al. were delivered earlier (kindergarten compared to primary school in our study), a different social skills program was used (i.e., a modified version of "I Can Problem Solve"), the social skills and parent programs were designed to fit together (rather than being two separate programs as in our case), and trained clinical psychologists delivered the parenting intervention (compared to parenting counselors in our case).

Second, although the participation rates in the parenting training were comparable to or better than those in other studies (e.g., Heinrichs et al., 2005; McTaggart and Sanders, 2003; Webster-Stratton et al., 2001), they were still limited. Thus, despite extensive efforts to increase program participation in our sample, it may be possible that parenting programs are not feasible for implementation among entire cohorts of children and parents.

Third, the quality of educational and behavioral programs offered in primary schools in Zurich is comparatively high, meaning that the lack of long-term intervention effects may be due to the high standards in the control condition. In a very comparable study to ours, namely the mentioned study by Lösel et al. (2013), slightly more reassuring long-term effects were found. In addition, some meta-analyses suggest that studies conducted in the United States tend to report higher average effect sizes than studies conducted in Europe (Piquero et al. 2009). These differences possibly reflect a better baseline provision with preventive strategies already embedded in the existing system.

Fourth, in most intervention studies, funding for examining long-term effects may be correlated with initial positive results. This may mean that the current picture on long-term effects may be contaminated by differential 'survival rates' of intervention studies. The present study, in contrast, is a comprehensive multi-informant longitudinal study in its own right and its long-term continuation was not dependent on treatment effects. It may therefore provide a more realistic estimate of treatment effects in field settings.

Fifth, there is convincing evidence that studies that involve researchers with a potential financial conflict of interests report higher effect sizes than independent field trials (Eisner 2009). Whether this is due to bias resulting from conflict of interests or better implementation quality is currently unclear. However, the findings reported here are in line with the review of the evidence by Petrosino and Soydan (2005).

Sixth, even though our data suggest that both the social skills program and the parent training were delivered to high standards (see Method section), it is still possible that the implementation quality was not optimal. For example, even though coaches assessed a part of the PATHS lessons, they were not present at all lessons, and even though the providers of the parenting training were certified Triple P therapists and estimated that 93% of the full course materials was delivered, these data were self-reported. Thus, the possibility that the programs were not implemented with fidelity cannot be dismissed.

Finally, although we had a high baseline participation rate, about one fifth of children did not participate in wave 1 of the study. It is therefore possible that inclusion of these children would have yielded different study results.

Notwithstanding these explanations, it is possible that, even under the best circumstances, it is simply not realistic to expect a long-term effect of these programs on later delinquency and problem behavior. Although prior analyses have shown that universal, school-based programs promote positive student behavior, the effects are small, even in the short-term, especially for conduct problems and antisocial behavior (Bradley et al. 2003; Durlak et al. 2013; Lösel et al. 2013). Compared to universal programs, programs that target at-risk groups yield larger effects (Lösel & Beelmann 2003; Wilson & Lipsey 2007). These findings are in line with research in other areas, such as the prevention of depression and substance use in youths, which has also shown limited effectiveness of universal approaches compared to selective and indicated interventions (Horowitz & Garber 2006; Tobler et al. 2000). Thus, although early universal prevention has become a highly fashionable strategy that is endorsed by many policy makers, the present findings suggest that even evidence-based programs may have no effects on later delinquency, substance use, or antisocial behavior.

References

Beelmann, A., & Lösel, F. (2006). Child social skills training in developmental crime prevention: Effects on antisocial behavior and social competence. Psicothema, 18, 603–610.

Bendixen, M., Endresen, I. M., & Olweus, D. (2003). Variety and frequency scales of antisocial involvement: Which one is better? Legal and Criminological Psychology, 8, 135–150.

Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., et al. (2010). The effects of a multiyear universal social--emotional learning program: The role of student and school characteristics. Journal of Consulting and Clinical Psychology, 78, 156.

Botvin, G. J., Baker, E., Dusenbury, L., Botvin, E. M., & Diaz, T. (1995). Long-term follow-up results of a randomized drug abuse prevention trial in a white middle-class population. Journal of the American Medical association, 273, 1106–1112.

Bradley, S. J., Jadaa, D. A., Brody, J., Landy, S., Tallett, S. E., Watson, W., et al., D. (2003). Brief psychoeducational parenting program: An evaluation and 1-year follow-up. Journal of the American Academy of Child & Adolescent Psychiatry, 42, 1171–1178.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). Applied multiple regression/correlation analysis for the behavioral sciences. (New York: Routledge)

Conduct Problems Prevention Research Group. (2010). Fast Track intervention effects on youth arrests and delinquency. Journal of Experimental Criminology, 6, 131–157.

Craig, W., Harel-Fisch, Y., Fogel-Grinvald, H., Dostaler, S., Hetland, J., Simons-Morton, B., et al. (2009). A cross-national profile of bullying and victimization among adolescents in 40 countries. International Journal of Public Health, 54, 216–224.

Crean, H. F., & Johnson, D. B. (2013). Promoting alternative thinking strategies (PATHS) and elementary school aged children's aggression: Results from a cluster randomized trial. American Journal of Community Psychology, 52, 56–72.

Currie, C., Zanotti, C., Morgan, A., Currie, D. De Looze, M., Roberts, C., et al. (Eds., 2012). Social determinants of health and well-being among young people. Health Behaviour in School-aged Children (HBSC) study: international report from the 2009/2010 survey. Copenhagen, WHO Regional Office for Europe, 2012 (Health Policy for Children and Adolescents, No. 6).

Dekovic, M., Slagt, M. I., Asscher, J.J., Boendermaker, L., Eichelsheim, V.I., & Prinzie, P. (2011). Effects of early prevention programs on adult criminal offending: A meta-analysis. Clinical Psychology Review, 31, 532–544

Dishion, T. J., Kavanagh, K., Schneiger, A., Nelson, S., & Kaufman, N. K. (2002). Preventing early adolescent substance use: A family-centered strategy for the public middle school. Prevention Science, 3, 191–201.

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool PATHS curriculum. The Journal of Primary Prevention, 28, 67–91.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. Child Development, 82, 405–432.

Eisner, M. (2009). No effects in independent prevention trials: Can we reject the cynical view? Journal of Experimental Criminology, 5, 163–183.

Eisner, M., Jünger, R., & Greenberg, M. (2006). Gewaltprävention durch die Förderung emotionaler und sozialer Kompetenzen in der Schule: Das PATHS/PFAD Curriculum. (Zürich: Universität Zürich)

Eisner, M.P., & Malti, T. (2015). Aggressive and violent behavior. (In M. E. Lamb (Vol. Ed.) & R.M. Lerner (Series Ed.), Handbook of child psychology and developmental science, Vol. 3: Social, emotional and personality development (7th edition, pp. 795–884). New York: Wiley.)

Eisner, M. P., Malti, T., & Ribeaud, D. (2011). Large-scale criminological field experiments. (In S. Gadd, S. Karstedt, & S. Messner (Eds.), Sage handbook of criminological research methods (pp. 410–424). London: Sage.)

Eisner, M., Manzoni, P., Ribeaud, D. & Schmid, R. (2003). Wirksame Gewaltprävention und -intervention bei Kindern und Jugendlichen in der Stadt Zürich.

Eisner, M., & Meidert, U. (2011). Stages of parental engagement in a universal parent training program. Journal of Primary Prevention, 32, 83–93.

Eisner, M., Nagin, D., Ribeaud, D., & Malti, T. (2012). Effects of a universal parenting program for highly adherent parents: a propensity score matching approach. Prevention Science, 13, 252–266.

Eisner, M., & Ribeaud, D. (2007). Conducting a criminological survey in a culturally diverse context. Lessons from the Zurich Project on the Social Development of Children. European Journal of Criminology, 4, 271–298.

Eisner, M., Ribeaud, D., Jünger, R., & Meidert, U. (2007). Frühprävention von Gewalt und Aggression: Ergebnisse des Zürcher Interventions- und Präventionsprojektes an Schulen [Early prevention of violence and aggression: Results of the Zurich Prevention and Intervention Project at schools. (Zürich: Rüegger)

Eisner, M., Ribeaud, D., & Topçuoglu, T. (2008). Indikatoren zur wirtschaftlichen, sozialen und kulturellen Lage von immigrierten Minderheiten in der Stadt Zürich. (Zürich: Stadt Zürich)

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. Educational and Psychological Measurement, 61, 713–740.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Structural Equation Modeling, 8, 430–457.

Enzmann, D., Marshall, I. H., Killias, M., Junger-Tas, J., Steketee, M., & Gruszczynska, B. (2010). Self-reported youth delinquency in Europe and beyond: First results of the Second International Self-Report Delinquency Study in the context of police and victimization data. European Journal of Criminology, 7, 159–183.

Farrington, D. P. (2012). Childhood risk factors for young adult offending. Onset and persistence. (In Lösel, F., Bottoms, A., & Farrington, D. P. (Eds.), Young adult offenders: Lost in transition? (pp. 48–64). Abingdon: Routledge.)

Farrington, D. P., & Welsh, B. C. (2003). Family-based prevention of offending: A meta-analysis. Australian & New Zealand Journal of Criminology, 36, 127–151.

Farrington, D. P., & Welsh, B. C. (2013). Randomized experiments in criminology: What has been learned from long-term follow-ups? (In B. C. Welsh, A. A. Braga, & G. J. N. Bruinsma (Eds.), Experimental criminology: Prospects for advancing science and public policy (pp. 111–140). New York: Cambridge University Press.)

Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S. M., & Donnelly, M. (2012). Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years. Campbell Systematic Reviews, 1–344.

Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. Social Science Research, 21, 1–56.

Greenberg, M. T., & Kusché, C. A. (2002). The PATHS curriculum: Follow up effects and mediational processes. Development and Psychopathology, 7, 1–36.

Greenberg, M. T., & Kusché, C. A. (1998). Preventive interventions for school-age deaf children: The PATHS curriculum. Journal of Deaf Studies and Deaf Education, 3, 49–63.

Greenberg, M. T., Mihalic, S. F., & Kusché, C. A. (1998). Promoting alternative thinking strategies (PATHS). (Boulder: Institute of Behavioral Science, University of Colorado)

Heinrichs, N., Bertram, H., Kuschel, A., & Hahlweg, K. (2005). Parent recruitment and retention in a universal prevention program for child behavior and emotional problems: Barriers to research and program participation. Prevention Science, 6, 275–286.

Heinrichs, N., Kliem, S., & Hahlweg, K. (2014). Four-year follow-up of a randomized controlled trial of Triple P group for parent and child outcomes. Prevention Science, 15, 233–245.

Horowitz, J. L., & Garber, J. (2006). The prevention of depressive symptoms in children and adolescents: A meta-analytic review. Journal of Consulting and Clinical Psychology, 74, 401–415.

Jaccard, J.,, & Turrisi, R. (2003). Interaction effects in multiple regression. (Thousand Oaks: Sage)

Jo, B. (2002). Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance. Statistics in Medicine, 21, 3161–3181.

Kam, C.-M., Greenberg, M. T., & Kusche, C. A. (2004). Sustained effects of the PATHS curriculum on the social and psychological adjustment of children in special education. Journal of Emotional and Behavioral Disorders, 12, 66–78.

Kam, C.-M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. Prevention Science, 4, 55–63.

Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. Journal of Abnormal Child Psychology, 36, 567–589.

Killias, M., Aebi, M., Herrmann, L., Dilitz, C., & Lucia, S. (2010). Switzerland. (In J. Junger-Tas, I. H. Marshall, D. Enzmann, M. Killias, M. Steketee, & B. Gruszczynska (Eds.), Juvenile delinquency in Europe and beyond. Results of the second International Self-Report Delinquency Study (pp. 79–95). Dordrecht: Springer.)

Koehler, J. A., Lösel, F., Akoensi, T. D., & Humphreys, D. K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. Journal of Experimental Criminology, 9, 19–43.

Kusché, C. A., & Greenberg, M. T. (1994). The PATHS curriculum. (Seattle: Developmental Research and Programs)

Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. Structural Equation Modeling: A Multidisciplinary Journal, 18, 649–662.

Lösel, F., & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized experiments. The Annals of the American Academy of Political and Social Science, 587, 84–109.

Lösel, F., Stemmler, M., & Bender, D. (2013). Long-term evaluation of a bimodal universal prevention program: effects on antisocial development from kindergarten to adolescence. Journal of Experimental Criminology, 9, 429–449.

Malti, T., Ribeaud, D., & Eisner, M. P. (2011). The effectiveness of two universal preventive interventions in reducing children's externalizing behavior: a cluster randomized controlled trial. Journal of Clinical Child & Adolescent Psychology, 40, 677–692.

Malti, T., Eisner, M.P., & Ribeaud, D. (2012). Effectiveness of a universal school-based social competence program: The role of child characteristics and economic factors. International Journal of Conflict and Violence, 6, 249–259.

Mason, W. A., Kosterman, R., Hawkins, J. D., Haggerty, K. P., & Spoth, R. L. (2003). Reducing adolescents' growth in substance use and delinquency: Randomized trial effects of a parent-training prevention intervention. Prevention Science, 4, 203–212.

McTaggart, P., & Sanders, M. R. (2003). The transition to school project: Results from the classroom. Australian e-Journal for the Advancement of Mental Health, 2, 1–12.

Montgomery, A. A., Peters, T. J., & Little, P. (2003). Design, analysis and presentation of factorial randomised controlled trials. *BMC Medical Research Methodology*, *3*, 26.

Muthén, L. K., & Muthén, B. O. (1998). Mplus user's guide. Seventh Edition. Los Angeles: Author. (Los Angeles: Muthén & Muthén) Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Mplus+user+guide#8

Nowak, C., & Heinrichs, N. (2008). A comprehensive meta-analysis of Triple P-Positive Parenting Program using hierarchical linear modeling: Effectiveness and moderating variables. Clinical Child and Family Psychology Review, 11, 114–144.

Olweus, D (1993). Bullying at school. What we know and what we can do. (Oxford: Blackwell)

Olweus, D. (1999). Sweden. In P. K. Smith, Y. Morita, J. Junger-Tas, D. Olweus, R. Catalano, & P. Slee (Eds.), The nature of school bullying: A cross-national perspective (pp. 7–27). (New York: Routledge)

Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. Journal of Experimental Criminology, 1, 435–450.

Piquero, A. R., Farrington, D. P., Welsh, B. C., Tremblay, R., & Jennings, W. G. (2009). Effects of early family/parent training programs on antisocial behavior and delinquency. Journal of Experimental Criminology, 5, 83–120.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. Law and Human Behavior, 29, 615.

Riggs, N. R., Greenberg, M. T., Kusché, C. A., & Pentz, M. A. (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: Effects of the PATHS curriculum. Prevention Science, 7, 91–102.

Roberts, C., & Torgerson, D. J. (1999). Understanding controlled trials: baseline imbalance in randomised controlled trials. British Medical Journal, 319, 185.

Sanders, M. R. (1999). Triple P-Positive Parenting Program: Towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. Clinical Child and Family Psychology Review, 2, 71–90.

Sanders, M. R. (2012). Development, evaluation, and multinational dissemination of the Triple P-Positive Parenting Program. Annual Review of Clinical Psychology, 8, 345–379.

Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014). The Triple P-Positive Parenting Program: A systematic review and meta-analysis of a multi-level system of parenting support. Clinical Psychology Review, 34, 337–357.

Simons-Morton, B. G., Farhat, T., Ter Bogt, T. F., Hublet, A., Kuntsche, E., Gabhainn, S. N., et al. (2009). Gender specific trends in alcohol use: cross-cultural comparisons from 1998 to 2006 in 24 countries and regions. International Journal of Public Health, 54, 199–208.

Smedler, A. C., Hjern, A., Wiklund, S., Anttila, S., & Pettersson, A. (2014). Programs for Prevention of Externalizing Problems in Children: Limited Evidence for Effect Beyond 6 Months Post Intervention. Child & Youth Care Forum, 1–26. (New York: Springer)

Spoth, R. L., Redmond, C., Trudeau, L., & Shin, C. (2002). Longitudinal substance initiation outcomes for a universal preventive intervention combining family and school programs. Psychology of Addictive Behaviors, 16, 129–134.

Sweeten, G. (2012). Scaling criminal offending. Journal of Quantitative Criminology, 28, 533–557.

Tobler, N. S., Roona, M. R., Ochshorn, P., Marshall, D. G., Streke, A. V., & Stackpole, K. M. (2000). School-based adolescent drug prevention programs: 1998 meta-analysis. Journal of Primary Prevention, 20, 275–336.

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivée, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. Journal of Abnormal Child Psychology, 19, 285–300.

Vitaro, F., Brendgen, M., & Tremblay, R. E. (2001). Preventive intervention: Assessing its effects on the trajectories of delinquency and testing for mediational processes. Applied Developmental Science, 5, 201–213.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. Journal of Clinical Child Psychology, 30, 283–302.

Wetzels, P., Enzmann, D., Mecklenburg, E., & Pfeiffer, C. (2001). Jugend und Gewalt. Eine repräsentative Dunkelfeldanalyse in München und acht anderen deutschen Städten [Youth and violence. A representative dark number study in Munich and eight other German cities]. (Baden-Baden: Nomos)

Wills, T. A., Sandy, J. M., Yaeger, A. M., Cleary, S. D., & Shinar, O. (2001). Coping dimensions, life stress, and adolescent substance use: a latent growth analysis. Journal of Abnormal Psychology, 110, 309.

Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. American Journal of Preventive Medicine, 33, S130–S143.

Wilson, P., Rush, R., Hussey, S., Puckering, C., Sim, F., Allely, C. S., … Gillberg, C. (2012). How evidence-based is an'evidence-based parenting program'? A PRISMA systematic review and meta-analysis of Triple P. BMC Medicine, 10, 130.

Windle, M., & Wiesner, M. (2004). Trajectories of marijuana use from adolescence to young adulthood: Predictors and outcomes. Development and Psychopathology, 16, 1007–1027.

Table 1. Means (standard errors) of outcome variables.

Variable Description				Age 13				Age 15			
Youth reports		Age 13	Age 15	Control	PATHS	Triple P	PATHS+ Triple P	Control	PATHS	Triple P	PATHS+ Triple P
Delinquency	M	1.03	1.18	1.06	1.12	.92	.94	1.04	1.05	1.13	1.08
	SE	.05	.05	.12	.11	.09	.09	.09	.10	.10	.11
Police contact	M	.07	.10	.08	.07	.07	.06	.09	.09	.09	.11
	SE	.01	.01	.02	.01	.01	.01	.02	.02	.02	.02
Substance use	M	.78	1.97	.74	.75	.77	.80	1.96	1.93	1.84	1.86
	SE	.03	.04	.07	.06	.07	.07	.09	.08	.09	.09
Peer aggression	M	1.72	1.77	1.68	1.75	1.68	1.74	1.75	1.69	1.70	1.78
	SE	.02	.02	.05	.04	.04	.05	.04	.04	.04	.05
Aggressive behavior	M	1.75	1.69	1.72	1.76	1.69	1.79	1.62	1.69	1.65	1.72
	SE	.02	.01	.03	.03	.03	.04	.03	.03	.03	.04
Prosocial behavior	M	3.56	3.6	3.61	3.55	3.59	3.51	3.64	3.63	3.59	3.57
	SE	.02	.02	.04	.04	.04	.04	.04	.03	.04	.04
Aggressive conflict resolution	M	1.65	1.62	1.59	1.69	1.60	1.62	1.53	1.58	1.63	1.64
	SE	.02	.02	.04	.04	.03	.04	.03	.03	.03	.04
Competent conflict resolution	M	3.23	3.35	3.30	3.20	3.27	3.22	3.36	3.43	3.32	3.27
	SE	.02	.02	.05	.05	.05	.05	.05	.05	.05	.05
Teacher reports											
Substance Use	M	1.02	1.09	1.01	1.01	1.01	1.01	1.10	1.07	1.08	1.09
	SE	<.01	.01	<.01	.01	<.01	.01	.02	.01	.01	.01
Delinquency	M	1.03	1.08	1.02	1.03	1.02	1.02	1.09	1.07	1.07	1.07
	SE	<.01	<.01	<.01	.01	<.01	.01	.01	.01	.01	.01
Aggressive behavior	M	.36	.35	.32	.33	.32	.35	.38	.33	.30	.31
	SE	.02	.01	.03	.03	.03	.03	.04	.03	.03	.03
Prosocial behavior	M	2.07	2.06	2.10	2.13	2.08	2.04	2.06	2.15	2.07	1.99
	SE	.02	.02	.05	.05	.05	.05	.05	.04	.05	.05
Non-aggressive Conduct Disorder	M	.23	.28	.20	.22	.20	.22	.31	.23	.25	.28
	SE	.01	.01	.02	.02	.02	.03	.03	.02	.03	.03

Table 2. Multilevel regression models (n = 1,580).

	Age 13 PATHS b [95%CI]	ES	Age 13 Triple P b [95%CI]	ES	Age 15 PATHS b [95%CI]	ES	Age 15 Triple P b [95%CI]	ES
Youth reports								
Delinquency	-.028 [-.176, .119]	-.016	-.15 [-.299, -.001]	-.083	-.04 [-.190, .109]	-.022	.004 [-.145, .153]	.002
Police contact*	.743 [.549, 1.007]	**-.157**	.772 [.568, 1.050]	**-.137**	1.103 [.840, 1.449]	.054	.968 [.714, 1.312]	-.018
Substance use*	.958 [.770, 1.191]	-.023	1.031 [.833, 1.276]	.016	1.086 [.905, 1.303]	.044	1.046 [.879, 1.245]	.024
Peer aggression	.043 [-.026, .112]	.057	.016 [-.051, .084]	.022	.041 [-.026, .108]	.054	-.008 [-.076, .060]	-.011
Aggressive behavior	.045 [-.002, .092]	.077	-.045 [-.090, .000]	-.077	.045 [.006, .084]	.081	-.014 [-.057, .030]	-.024
Prosocial behavior	-.065 [-.116, -.014]	-.096	-.006 [-.059, .047]	-.009	-.011 [-.064, .041]	-.018	-.046 [-.096, .005]	-.072
Aggressive conflict resolution	.038 [-.023, .098]	.057	-.044 [-.105, .017]	-.067	.017 [-.036, .071]	.029	.036 [-.017, .088]	.059
Competent conflict resolution	-.075 [-.138, -.011]	-.087	.008 [-.059, .075]	.009	.051 [-.013, .114]	.063	-.08 [-.141, -.019]	**-.100**
Teacher reports								
Substance use	-.003 [-.012, .005]	-.036	-.005 [-.013, .004]	-.049	-.016 [-.039, .006]	-.070	-.011 [-.033, .010]	-.048
Deviance	0 [-.007, .008]	.005	-.005 [-.013, .003]	-.060	-.01 [-.024, .005]	-.065	-.016 [-.030, -.002]	**-.109**
Aggressive behavior	.029 [-.018, .075]	.054	-.013 [-.060, .034]	-.024	-.039 [-.096, .019]	-.077	-.075 [-.132, -.018]	**-.149**
Prosocial behavior	-.027 [-.124, .071]	-.032	-.06 [-.149, .029]	-.073	-.028 [-.120, .064]	-.036	-.045 [-.138, .047]	-.058
Non-aggressive Conduct Disorder	.017 [-.019, .054]	.040	-.011 [-.049, .027]	-.025	-.031 [-.095, .034]	-.064	-.023 [-.085, .039]	-.048

Note: Estimates adjusted for clustering within 56 schools and baseline control variables.
* = logistic regression models with Odds Ratio and confidence interval reported in the square brackets.
Estimates in **bold** statistically significant at $p < .05$.
ES = Cohen's d effect size. In the logistic models, the ES was computed given the variance of the modelled continuous response variable underlying the ordered-categorical outcome.

Table 3. Multilevel regression models using strict inclusion criteria (i.e., participation at age 7 and either age 13 or age 15, n = 1,275).

	Age 13 PATHS B (CI)	ES	Age 13 Triple P B (CI)	ES	Age 15 PATHS B (CI)	ES	Age 15 Triple P B (CI)	ES
Youth reports								
Delinquency	-.03 [-.193, .132]	-.017	-.161 [-.324, .002]	-.090	-.104 [-.260, .052]	-.056	-.031 [-.190, .129]	-.017
Police contact*	.652 [.464, .916]	**-.225**	.791 [.565, 1.108]	**-.123**	.953 [.722, 1.259]	-.026	1.039 [.762, 1.416]	.021
Substance use*	.98 [-.264, .223]	-.011	.985 [-.254, .224]	-.008	1.04 [.870, 1.245]	.021	.984 [.820, 1.181]	-.008
Peer aggression	.034 [-.041, .108]	.043	.016 [-.057, .090]	.021	.013 [-.052, .079]	.018	-.02 [-.088, .047]	-.027
Aggressive behavior	.095 [.009, .103]	.095	-.083 [-.098, .000]	-.083	.036 [-.004, .077]	.067	-.013 [-.058, .032]	-.024
Prosocial behavior	-.043 [-.096, .009]	-.064	-.027 [-.083, .029]	-.040	.007 [-.044, .058]	.011	-.065 [-.118, -.012]	**-.104**
Aggressive conflict resolution	.028 [-.034, .089]	.042	-.034 [-.095, .028]	-.051	-.009 [-.058, .041]	-.015	.031 [-.020, .082]	.053
Competent conflict resolution	-.103 [-.174, -.031]	**-.121**	-.021 [-.099, .057]	-.025	.055 [-.012, .122]	.069	-.093 [-.156, -.030]	**-.117**
Teacher reports								
Substance use	-.005 [-.015, .004]	-.057	-.002 [-.011, .007]	-.023	-.023 [-.048, .002]	-.096	-.019 [-.043, .005]	-.080
Deviance	-.001 [-.008, .006]	-.018	-.003 [-.011, .004]	-.038	-.015 [-.031, .000]	-.101	-.021 [-.036, -.006]	**-.138**
Aggressive behavior	.025 [-.020, .069]	.046	-.009 [-.059, .041]	-.017	-.051 [-.109, .007]	-.102	-.068 [-.126, -.009]	**-.135**
Prosocial behavior	-.019 [-.122, .083]	-.024	-.089 [-.185, .008]	-.108	-.014 [-.113, .085]	-.018	-.08 [-.179, .019]	-.103
Non-Aggressive Conduct Disorder	.025 [-.011, .062]	.057	-.009 [-.048, .030]	-.020	-.047 [-.112, .019]	-.096	-.011 [-.074, .052]	-.023

Note: Estimates adjusted for clustering within 56 schools and baseline control variables.
* = logistic regression models with Odds Ratio and confidence interval reported in the square brackets.
Estimates in **bold** statistically significant at $p < .05$.
ES = Cohen's d effect size. In the logistic models, the ES was computed given the variance of the modelled continuous response variable underlying the ordered-categorical outcome.

Table 4. Multilevel regression models with interaction term for the combined treatment (PATHS X Triple P; only significant results included).

	Age 13					
	PATHS		Triple P		PATHS x Triple P	
	b [95%CI]	ES	b [95%CI]	ES	b [95%CI]	ES
Competent conflict resolution	-.206 [-.311, -.101]	**-.243**	-.127 [-.210, -.044]	**-.149**	.22 [.083, .356]	**.259**

Note: Estimates adjusted for clustering within 56 schools and baseline control variables.

ES = Cohen's d effect size. In the logistic models, the ES was computed given the variance of the modelled continuous response variable underlying the ordered-categorical outcome.
Only significant results displayed, i.e., for models using strict inclusion criteria only (i.e., participation at age 7 and either age 13 or age 15, n = 1,275).

Figure Captions.

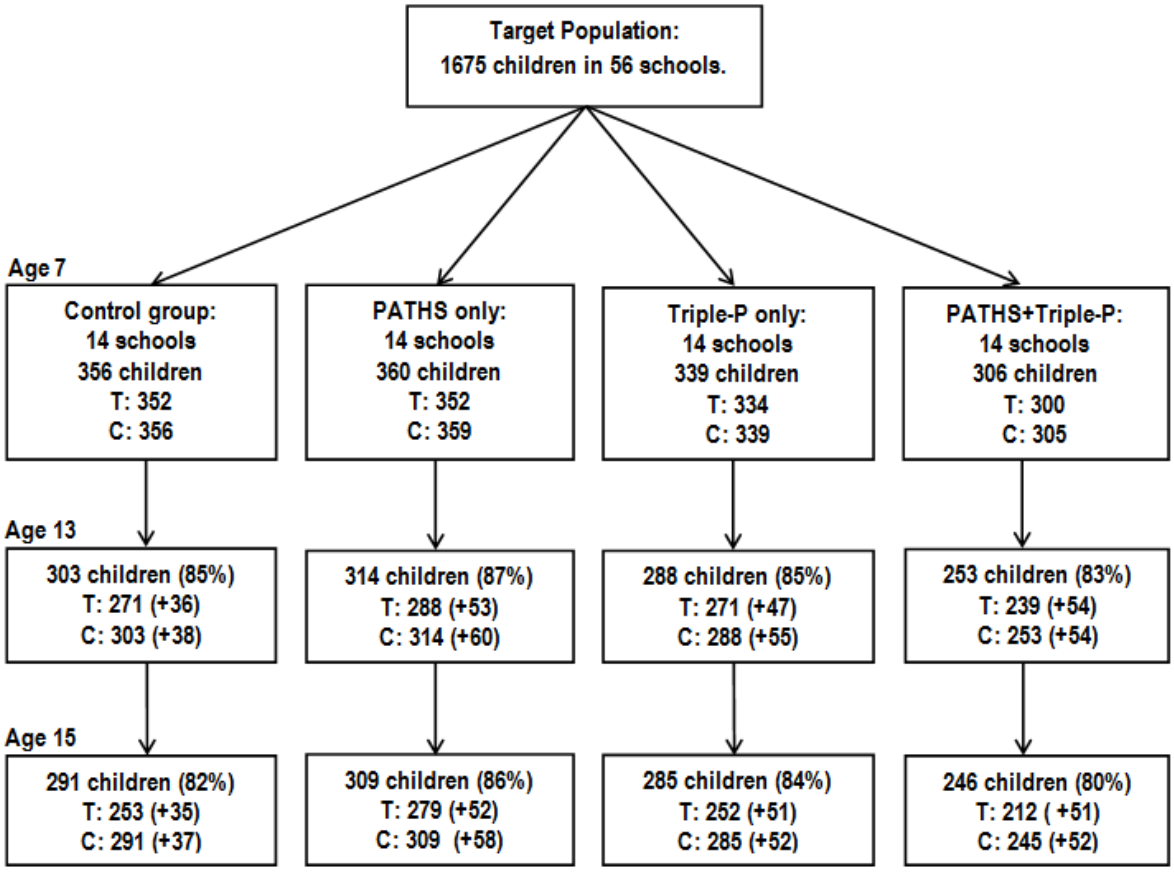

Figure 1. Flow diagram of study participation and treatment conditions.

T = Teacher-reported data. C = Child-reported data. Participants who did not take part in wave 1 in parentheses (+)

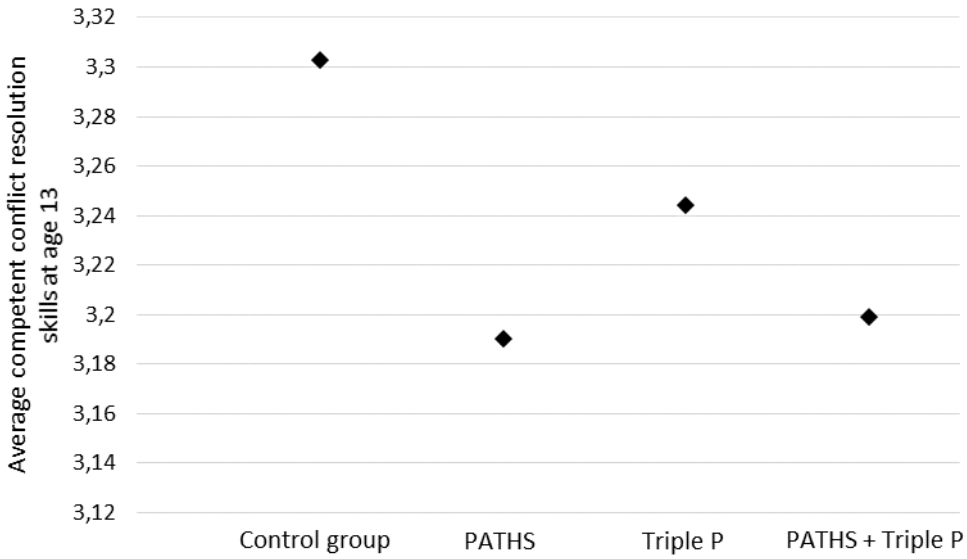

Figure 2. Competent conflict resolution at age 13 by treatment group.

Note: Estimates adjusted for clustering within 56 schools and baseline control variables.